

AFRL-VA-WP-TP-2003-318

**ENCODING STRATEGY FOR
MAXIMUM NOISE TOLERANCE
BIDIRECTIONAL ASSOCIATIVE
MEMORY**

**Dan Shen
Jose B. Cruz, Jr.**



JULY 2003

Approved for public release; distribution is unlimited.

© 2003 IEEE

This work is copyrighted. The United States has for itself and others acting on its behalf an unlimited, paid-up, nonexclusive, irrevocable worldwide license. Any other form of use is subject to copyright restrictions.

**AIR VEHICLES DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7542**

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YY) July 2003		2. REPORT TYPE Journal Article Preprint		3. DATES COVERED (From - To)		
4. TITLE AND SUBTITLE ENCODING STRATEGY FOR MAXIMUM NOISE TOLERANCE BIDIRECTIONAL ASSOCIATIVE MEMORY				5a. CONTRACT NUMBER F33615-01-C-3151		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER 69199F		
6. AUTHOR(S) Dan Shen Jose B. Cruz, Jr.				5d. PROJECT NUMBER ARPF		
				5e. TASK NUMBER 04		
				5f. WORK UNIT NUMBER T2		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Ohio State University Columbus, OH 43210				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Vehicles Directorate Air Force Research Laboratory Air Force Materiel Command Wright-Patterson Air Force Base, OH 45433-7542				10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/VAC		
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-VA-WP-TP-2003-318		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.						
13. SUPPLEMENTARY NOTES Journal article preprint submitted in June 2003 for publication in <i>IEEE Transactions on Neural Networks</i> . Technical paper contains color. © 2003 IEEE. This work is copyrighted. The United States has for itself and others acting on its behalf an unlimited, paid-up, nonexclusive, irrevocable worldwide license. Any other form of use is subject to copyright restrictions.						
14. ABSTRACT In this paper, the Basic Bidirectional Associative Memory (BAM) is extended by choosing weights in the correlation matrix, for a given set of training pairs, which result in a maximum noise tolerance set for BAM. This optimized BAM will recall the correct training pair if an input pair is within the maximum noise tolerance set. We define a hyper-radius, and we prove that for a given set of training pairs, the maximum noise tolerance set is the largest, in the sense that at least one pair outside the maximum noise tolerance set, and within a Hamming distance one larger than the hyper-radius associated with the maximum noise tolerance set, will not converge to the correct training pair. A standard Genetic Algorithm (GA) is used to calculate the weights to maximize the objective function which generates a maximum tolerance set for BAM. Computer simulations are presented to illustrate the error correction and fault tolerance properties of the optimized BAM.						
15. SUBJECT TERMS BAM, energy well hyper-radius, GA, tolerance set, training set						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 22	19a. NAME OF RESPONSIBLE PERSON (Monitor) Mark J. Mears 19b. TELEPHONE NUMBER (Include Area Code) (937) 255-8685	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified				

Encoding Strategy For Maximum Noise Tolerance Bidirectional Associative Memory

Dan Shen and Jose B. Cruz, Jr., *Life Fellow, IEEE*

Abstract

In this paper, the Basic Bidirectional Associative Memory (BAM) is extended by choosing weights in the correlation matrix, for a given set of training pairs, which result in a maximum noise tolerance set for BAM. This optimized BAM will recall the correct training pair if an input pair is within the maximum noise tolerance set. We define a hyper-radius, and we prove that for a given set of training pairs, the maximum noise tolerance set is the largest, in the sense that at least one pair outside the maximum noise tolerance set, and within a Hamming distance one larger than the hyper-radius associated with the maximum noise tolerance set, will not converge to the correct training pair. A standard Genetic Algorithm (GA) is used to calculate the weights to maximize the objective function which generates a maximum tolerance set for BAM. Computer simulations are presented to illustrate the error correction and fault tolerance properties of the optimized BAM.

Index Terms

BAM, Energy Well Hyper-Radius, GA, Tolerance Set, Training Set

Manuscript received June 2, 2003. This work was sponsored by the Defense Advanced Research Project Agency (DARPA) under Contract F33615-01-C3151 issued by the AFRL/VAK.

D. Shen is with The Ohio State University, Columbus, OH 43210 USA(email: shen.100@osu.edu).

J. B. Cruz, Jr. is with The Ohio State University, Columbus, OH 43210 USA(email: jbcruz@ieee.org).

I. INTRODUCTION

In 1968, Anderson [6] proposed a memory structure named Linear Associative Memory (LAM), which can be used in hetero-associative pattern recognition. Since LAM is noise sensitive, Optimal Linear Associative Memory was introduced by Wee [7] and Kohonen [8], which extended the LAM by absorbing the noise. Although good results can be obtained using these early approaches, many theoretical and practical issues such as network stability and storage capacity were still unresolved. In 1988, Kosko [1] presented the theory of bidirectional associative memory by generalizing the Hopfield network model.

As a class of artificial neural networks, Bidirectional Associative Memories (BAM) provide massive parallelism, high error correction and high fault tolerance ability. However, to form a good BAM, a good encoding strategy was required. This field has received extensive attention from researchers and a substantial effort has been devoted to various learning rules. Kosko [1] has provided a correlation learning strategy and proved that the BAM process will converge after a finite number of interactions. However, the correlation matrix used by Kosko cannot guarantee that the energy of any training pair is a local minimum. That is, it can not guarantee recall of any training pair even for a very small set of training data.

During the following years, various encoding strategies and learning rules were proposed to improve the capacity and the performance of BAM. In 1990, Wang, Cruz, and Mulligan [2] introduced two BAM encoding schemes to increase the recall performance with a trade off of more neurons. These are multiple training methods, which guarantee the recall of all training pairs [3]. In 1993 and 1994, Leung [9] [10] present the Enhanced Householder Encoding Algorithm (EHCA), which was improved by Lenze [11] in 2001, to enlarge the capacity. In 1995, Wang and Don [12] introduced the exponential bidirectional associative memory (eBAM), which uses an exponential encoding rule rather than the correlation scheme.

However, these methods have focused on the training set or capacity only. The noisy neighbor pairs and the noise tolerance set of BAM have been ignored. In this paper, we are especially interested in the approach proposed by Wang, Cruz, and Mulligan [2] [3] and extend the BAM by choosing the weights for training pairs in the BAM correlation matrix, which can maximize the noise tolerance set, for a given set of training pairs, such that any noisy input pair within the tolerance set will converge to the correct training pair.

Some basic concepts of BAM are reviewed in Section II. Then, the multiple training concept is extended in Section III with the optimization-based encoding strategy for constructing the correlation matrix. Two lemmas and a theorem about the new encoding rule are proved in the same section. These provide the foundation for constructing the maximum noise tolerance set. We present a numerical example in Section IV to illustrate the effectiveness of the extended BAM. In this example, a standard GA is used to resolve the nonlinear optimal problem and obtain the optimum training weights. Finally, we draw conclusions and enumerate some possible future extensions in Section V.

II. BIDIRECTIONAL ASSOCIATIVE MEMORY

BAM is a two-layer hetero-associative feedback neural network model first introduced by Kosko [1]. As shown in Fig. 1, the input layer L_A includes n binary valued neurons (a_1, a_2, \dots, a_n) and the output layer L_B comprises m binary valued components (b_1, b_2, \dots, b_m). Now we have $L_A = \{0,1\}^n$ and $L_B = \{0,1\}^m$. BAM can be denoted as a bi-directional mapping in vector space $M : R_n \leftrightarrow R_m$. The training pairs can be stored in the correlation matrix as follows:

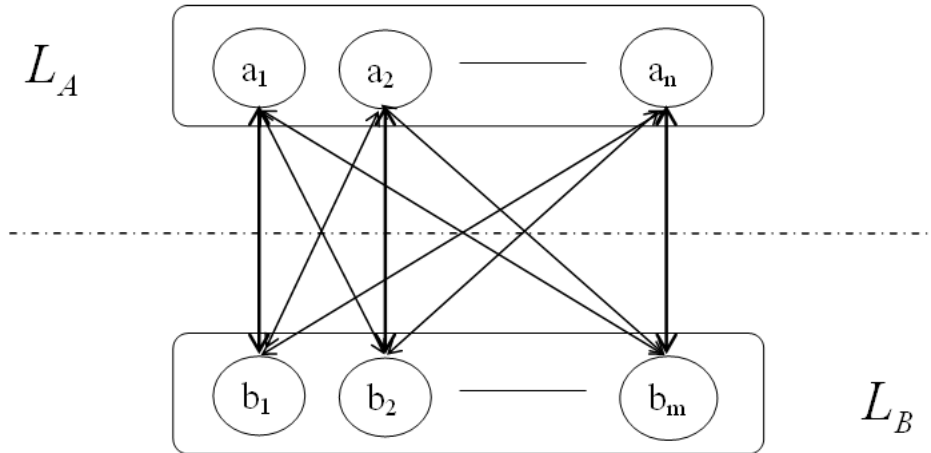


Fig. 1. Structure of Bidirectional Associative Memory

$$M = \sum_{i=1}^N X_i^T Y_i$$

where X_i and Y_i are the bipolar mode of A_i and B_i respectively, i.e.

$$\begin{cases} X_i &= 2A_i - \vec{1} \\ Y_i &= 2B_i - \vec{1} \end{cases}$$

If inputs X_1, X_2, \dots, X_N are orthogonal to each other, i.e.

$$X_i X_j^\top = \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j \end{cases}$$

then,

$$X_i M = X_i \left(\sum_{j=1}^N X_j^\top Y_j \right) = X_i X_i^\top Y_i + \sum_{j=1, j \neq i}^N X_i X_j^\top Y_j = Y_i$$

To obtain higher accuracy for associative memory and retrieve one of the nearest training inputs, the output Y can be fed back to BAM. Starting with a pair (α_0, β_0) , determine a sequence $(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots$, until it finally converges to an equilibrium point (α_F, β_F) . If BAM converges for every training pair, M is said to be bidirectional stable.

The sequence can be obtained as follows:

$$[\beta_{i+1}]_k = \begin{cases} 1 & , \quad [\alpha_i M]_k > \varepsilon_k \\ [\beta_i]_k & , \quad [\alpha_i M]_k = \varepsilon_k \\ -1 & , \quad [\alpha_i M]_k < \varepsilon_k \end{cases} \quad [\alpha_{i+1}]_k = \begin{cases} 1 & , \quad [\beta_i M]_k > \delta_k \\ [\alpha_i]_k & , \quad [\beta_i M]_k = \delta_k \\ -1 & , \quad [\beta_i M]_k < \delta_k \end{cases}$$

where $[\bullet]_k$ is the k_{th} element of the vector. ε_k and δ_k are two thresholds for the k_{th} element of α_i and β_i respectively. If $(\varepsilon, \delta)^\top = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N, \delta_1, \delta_2, \dots, \delta_N)^\top = \vec{0}$, then this kind of BAM is called homogeneous. Others are called non-homogeneous BAM.

For each pair, the Lyapunov or energy function is defined as,

$$E = \begin{cases} -\alpha M \beta^\top & , \quad (\varepsilon, \delta)^\top = \vec{0} \\ -\alpha M \beta^\top + \alpha \varepsilon^\top + \beta \delta^\top & , \quad (\varepsilon, \delta)^\top \neq \vec{0} \end{cases}$$

Kosko [1] and Haines *et al.* [4] have proved that after a finite number of iterations, E converges to a local minimum, where the corresponding pair (α_F, β_F) is a stable point.

McEliece *et al.* [5] have shown that if the training pairs are even coded (± 1 with probability 0.5) and n -dimensional, the storage capacity of the homogeneous BAM is $\frac{n}{2 \log_2 n}$. That means, if L even-coded stable states are chosen uniformly at random, the maximum value of L in order that most of the L original vectors are accurately recalled is $\frac{n}{2 \log_2 n}$.

For the non-homogeneous BAM, Haines and Hecht-Nielsen [4] have pointed out that the possible number of the stable states is between 1 and $2^{\min(m,n)}$. However, since these stable states are chosen in a rigid geometrical procedure, the storage capacity of the non-homogeneous BAM is less than the maximum number. Haines and Hecht-Nielsen [4] also have shown that for N same dimensional and uniformly randomly chosen training pairs with $(4 + \log_2^n)$ exactly entries equal to $+1$ and $(n - 4 - \log_2^n)$ entries equal to -1 , if $N < \frac{0.68n^2}{[\log_2^n + 4]^2}$, then a non-homogeneous BAM can be constructed so that approximately 98% of these chosen pairs can be stable states.

III. ENCODING STRATEGY FOR BAM WITH MAXIMUM NOISE TOLERANCE SET

In this new enhanced model, we start with a weighted learning rule of BAM similar to the Multiple Training Strategy in [3]. For a given set of training pairs $\{(X_i, Y_i)\}_{i=1}^N$, the weighted correlation matrix is

$$M = \sum_{i=1}^N w_i X_i^\top Y_i \quad (1)$$

where,

$$\begin{aligned} X_i &= (x_{i1}, x_{i2}, \dots, x_{iQ}) \\ Y_i &= (y_{i1}, y_{i2}, \dots, y_{iP}) \end{aligned}$$

Q and P are the lengths of the input and output patterns respectively. $W = (w_1, w_2, \dots, w_N)$ is the vector of training weights. In [3], necessary and sufficient conditions are derived for choosing W such that each training pair can be recalled correctly.

The energy of a training pair (X_i, Y_i) is defined as

$$E(X_i, Y_i, M) = -X_i M Y_i^\top \quad (2)$$

If the energy of one training pair is lower than all its neighbors with one Hamming distance away from it, then the training pair can be recalled correctly.

The neighbor pairs with $n \in I$ Hamming distance away from a pair (X_i, Y_i) is defined as

$$\Omega(X_i, Y_i, n) = \begin{cases} \{(X, Y) | H_x(X_i, X) + H_y(Y_i, Y) = n\} & , \quad n > 0 \\ (X_i, Y_i) & , \quad n \leq 0 \end{cases}$$

where $H_x(X_i, X)$ is the Hamming distance between layers X_i and X , and $H_y(y_i, y)$ is the Hamming distance between layers Y_i and Y .

Lemma 1: If a training weight vector $W = [w_1, w_2, \dots, w_n]^\top$ satisfies

$$[\Gamma_1, \Gamma_2, \dots, \Gamma_N]^\top W \geq 0 \quad (3)$$

where,

$$\Gamma_i = \begin{bmatrix} \eta_{i1}^{A1} & \dots & \eta_{iN}^{A1} \\ \vdots & & \vdots \\ \eta_{i1}^{AQ} & \dots & \eta_{iN}^{AQ} \\ \eta_{i1}^{B1} & \dots & \eta_{iN}^{B1} \\ \vdots & & \vdots \\ \eta_{i1}^{BP} & \dots & \eta_{iN}^{BP} \end{bmatrix} \quad \begin{aligned} \eta_{ij}^{Ak} &= A_i X_j^\top Y_j B_i^\top - A_i^k X_j^\top Y_j B_i^\top \\ \eta_{ij}^{Bk} &= A_i X_j^\top Y_j B_i^\top - A_i X_j^\top Y_j (B_i^k)^\top \\ A_i^k(B_i^k) &\text{ differs from } A_i(B_i) \text{ only in the } k\text{-th bit} \end{aligned}$$

Then, $\exists \Psi \in I^+$, such that any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $1 \leq n \leq \Psi$ has higher energy than any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$.

Proof: Wang, Cruz, and Mulligan [2] have proved that if a training weight vector W satisfies condition (3), then all training pairs can be recalled correctly. Since a training pair P_i can be recalled correctly if and only if P_i is a local minimum on the energy surface, any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, 1)$ has higher energy than any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, 0) \right]$. So, at least $\exists \Psi = 1$ satisfying that any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $1 \leq n \leq \Psi$ has higher energy than any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$. ■

Definition 1: For a BAM(W, M) satisfying condition (3), we define the maximum Ψ as the energy well hyper-radius F which satisfies the following:

- 1) $F \in I^+$
- 2) any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $n \in I$ and $1 \leq n \leq F$ has higher energy than any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$;
- 3) at least one pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, F+1)$ has energy lower than or equal to that of at least one pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, F) \right]$.

Lemma 2: Given a desired training pair set $\{(X_i, Y_i)\}_{i=1}^N$, a weight vector W satisfying condition (3), for the associated energy well hyper-radius F , if we define $V_i(F-1, M) =$

$\{(X, Y) | H_x(X, X_i) + H_y(Y, Y_i) \leq F - 1\}$ for each i , $1 \leq i \leq N$, then,

- 1) any input pair in the set $V_i(F - 1, M)$ converges to the training pair (X_i, Y_i) ;
- 2) for any i and j such that $1 \leq i \neq j \leq N$, we have $V_i(F - 1, M) \cap V_j(F - 1, M) = \emptyset$;
- 3) an upper bound of the energy well hyper-radius $F(M)$ is

$$\hat{F} = \left\lfloor \frac{1}{2} \min \left(\min_{0 \leq i \neq j \leq N} H_x(X_i, X), \min_{0 \leq i \neq j \leq N} H_y(Y_i, Y) \right) + 1 \right\rfloor$$

Proof: From *Lemma 1* and *Definition 1*, since W satisfying (3), its associated energy well hyper-radius $F \geq 1$.

1) Kosko [1] has pointed out that when a pair is input to a BAM, the network quickly evolves to a system energy local minimum. For any input pair in $V_i(F - 1, M)$, there is a high energy "hill" around it. So it is guaranteed that BAM evolves to some pair $(X, Y) \in V_i(F - 1, M)$. Since (X_i, Y_i) is the only system energy local minimum, any input pair in the set $V_i(F - 1, M)$ converges to the training pair (X_i, Y_i) .

2) For any $1 \leq i \neq j \leq N$, if $V_i(F - 1, M) \cap V_j(F - 1, M) \neq \emptyset$, then there is at least one pair $(X, Y) \in V_i(F - 1, M) \cap V_j(F - 1, M)$. From conclusion 1) which we have just proved, (X, Y) converges to the training pair (X_i, Y_i) and (X_j, Y_j) . It implies that $(X_i, Y_i) \equiv (X_j, Y_j)$ which is inconsistent with the condition that $i \neq j$. So, for any i and j such that $1 \leq i \neq j \leq N$, $V_i(F - 1, M) \cap V_j(F - 1, M) = \emptyset$.

3) From the conclusion 2) that for any i and any j , $1 \leq i \neq j \leq N$, we have $V_i(F - 1, M) \cap V_j(F - 1, M) = \emptyset$, then we obtain $F - 1 \leq \frac{1}{2} \min(H_x(X_i, X), H_y(Y_i, Y))$, so an upper bound for the energy well hyper-radius is

$$\hat{F} = \left\lfloor \frac{1}{2} \min \left(\min_{0 \leq i \neq j \leq N} H_x(X_i, X), \min_{0 \leq i \neq j \leq N} H_y(Y_i, Y) \right) + 1 \right\rfloor$$

■

Definition 2: For a given training pair set $\{(X_i, Y_i)\}_{i=1}^N$ with a weight vector W and the associated energy well hyper-radius $F \geq 1$, we define $V(M) = \bigcup_{i=1}^N V_i(F - 1, M)$ as the *noise tolerance set* of $\text{BAM}(W, M)$.

Any pair in $V(M)$ input to $\text{BAM}(W, M)$ converges to the correct training pair.

We want to find the optimal training weight vector W^* which can generate a correlation matrix M^* with the maximum energy well hyper-radius F^* and the optimum noise tolerance

set $V^*(M^*) \supseteq \text{any } V(M)$. In [3], Wang *et al.* just considered neighbors with one Hamming distance, corresponding to $F = 1$, and $V(M) = \{(X_i, Y_i)\}_{i=1}^N$. Their method does not provide any information for determining a noise tolerance set $V(M) \supset \{(X_i, Y_i)\}_{i=1}^N$.

For each training pair (X_i, Y_i) in a training set $\{(X_i, Y_i)\}_{i=1}^N$ and M formed from the training set by equation (1), we define the energy of any neighbor

$$E_i^{m,p}(k_1, k_2, \dots, k_m; t_1, t_2, \dots, t_p; M) = -[X_i^m(k_1, k_2, \dots, k_m)]M[Y_i^p(t_1, t_2, \dots, t_p)]^\top \quad (4)$$

where,

$$(X_i^m(k_1, k_2, \dots, k_m), Y_i^p(t_1, t_2, \dots, t_p)) \in \Omega(X_i, Y_i, m + p).$$

(k_1, k_2, \dots, k_m) are the position indices that the m bits with the complementary values (in bipolar mode, the complementary value of -1(+1) is +1(-1); in binary mode, the complementary value of 1(0) is 0 (1)) for the input pattern X_i

$$1 \leq k_i \leq Q \quad \text{and} \quad k_i \neq k_j \text{ if } 1 \leq i \neq j \leq m \quad (5)$$

while (t_1, t_2, \dots, t_p) has a similar meaning for the output pattern Y_i

$$1 \leq t_i \leq P \quad \text{and} \quad t_i \neq t_j \text{ if } 1 \leq i \neq j \leq p \quad (6)$$

Also define

$$\begin{cases} X_i^0 &= X_i \\ Y_i^0 &= Y_i \\ E_i^{0,0} &= E(X_i, Y_i, M) \end{cases} \quad \phi(x) = \begin{cases} 1 &, x > 0 \\ 0 &, x \leq 0 \end{cases} \quad (7)$$

Then, for a fixed weight vector $W = (w_1, w_2, \dots, w_N)$, the object function is defined as

$$f(W) = \sum_{i=1}^N \bar{E}_i(M) \quad (8)$$

where $\bar{E}_i(M)$ is a weighted sum of energy difference between any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $1 \leq n \leq \hat{F}$ and any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$.

$$\bar{E}_i(M) = \sum_{m=0}^{\hat{F}} \sum_{p=\max(0, 1-m)}^{\hat{F}-m} \gamma_{m,p} \sum_{(5)} \sum_{(6)} \bar{E}_i^{m,p}(k_1, k_2, \dots, k_m; t_1, t_2, \dots, t_p; M) \quad (9)$$

where,

$\sum_{(5)} \sum_{(6)}$ means all combinations of k_1, k_2, \dots, k_m and t_1, t_2, \dots, t_p which satisfying condition (5)

and (6) respectively.

if $m \geq 2$ and $p \geq 2$ then,

$$\begin{aligned} \bar{E}_i^{m,p}(k_1, k_2, \dots, k_m; t_1, t_2, \dots, t_p; M) = \\ \prod_{(k'_1, k'_2, \dots, k'_{m-1}) \subset (k_1, k_2, \dots, k_m)} \phi \left(\frac{E_i^{m,p}(k_1, k_2, \dots, k_m; t_1, t_2, \dots, t_p; M) -}{E_i^{m-1,p}(k'_1, k'_2, \dots, k'_{m-1}; t_1, t_2, \dots, t_p; M)} \right) \times \\ \prod_{(t'_1, t'_2, \dots, t'_{p-1}) \subset (t_1, t_2, \dots, t_p)} \phi \left(\frac{E_i^{m,p}(k_1, k_2, \dots, k_m; t_1, t_2, \dots, t_p; M) -}{E_i^{m,p-1}(k_1, k_2, \dots, k_m; t'_1, t'_2, \dots, t'_{p-1}; M)} \right) \end{aligned} \quad (10)$$

if if $m = 1$ and $p = 1$, then,

$$\bar{E}_i^{1,1}(k_1; t_1; M) = \phi \left(E_i^{1,1}(k_1; t_1; M) + X_i M [Y_i^1(t_1)]^\top \right) \phi \left(E_i^{1,1}(k_1; t_1; M) + X_i^1(k_1) M Y_i^\top \right)$$

if if $m = 0$ and $p = 1$, then,

$$\bar{E}_i^{0,1}(t_1; M) = \phi \left(-X_i M [Y_i^1(t_1)]^\top - E_i^{0,0} \right)$$

if if $m = 1$ and $p = 0$, then,

$$\bar{E}_i^{1,0}(k_1; M) = \phi \left(-X_i^1(k_1) M Y_i^\top - E_i^{0,0} \right)$$

and

$$\gamma_{m,p}(x) = \begin{cases} 1 & , \quad x > 0 \\ -H_{m+p} & , \quad x \leq 0 \end{cases} \quad (11)$$

The series H_l can be generated by the following formula,

$$\begin{cases} H_{\hat{F}+1} = -1 \\ H_{\hat{F}} = 1 \\ H_{l-1} = N \sum_{i=l}^{\hat{F}} (H_i + 1) \binom{P+Q}{i} \quad , \quad l = \hat{F}, \hat{F} - 1, \dots, 2 \end{cases} \quad (12)$$

$$\text{where } \binom{n}{m} = \frac{m!}{m!(n-m)!} \text{ for any } n \geq m \geq 0, n \in I, m \in I$$

It is obvious that series H_l is strictly decreasing.

Maximum Noise Tolerance Theorem: Given a set of training pairs $\{(X_i, Y_i)\}_{i=1}^N$ and at least one W satisfying the condition of *Lemma 1*, and if W^* denotes the W that maximizes $f(W)$, where f is given in (4) - (12),

$$W^* = \arg \max_W f(W) \quad (13)$$

then,

1) The $\text{BAM}(W^*, M^*)$ has the maximum energy well hyper-radius $1 \leq F^* = r \leq \hat{F}$, where r uniquely satisfies,

$$N \sum_{i=1}^r \binom{Q+p}{i} - N \sum_{j=r+1}^{\hat{F}} H_j \binom{Q+P}{j} \leq f(W^*) \leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r+1} \quad (14)$$

2) $V^*(M^*) = \bigcup_{i=1}^N V_i(F^* - 1, M^*) \supseteq \text{any } V(M)$, i.e. for any $F' > F^*$, there is at least one pair $(X', Y') \in \bigcup_{i=1}^N V_i(F' - 1, M)$ such that if it is input to BAM, the output layer will not converge to the correct training pair.

Proof: We divide the proof into three parts. The first one is to show that r uniquely satisfies inequality (14). The second is to prove that $F^* = r$ is the maximum energy well hyper-radius. The last one is to show that $V^*(M^*) = \bigcup_{i=1}^N V_i(F^* - 1, M^*) \supseteq \text{any } V(M)$.

Firstly, given a training weight vector W and energy well hyper-radius F , $f(W)$ depends on the training pair set $\{(X_i, Y_i)\}_{i=1}^N$. Since for any pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $n \geq 1$ we put a penalty value $-H_n$ on the object function if (X, Y) has energy lower than or equal to that of any neighbor pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$ and is H_l a strictly decreasing series, the object function $f(W)$ takes the largest value when only one neighbor pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, F+1)$ has energy lower than or equal to that of one pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, F) \right]$. On the other hand, when any neighbor pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $n \geq F+1$ has energy lower than or equal to that of any pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n) \right]$, $f(W)$ takes the lowest value. So, inequality (14) holds.

It can be shown by contradiction that only one unique r satisfies the inequality (14).

If there is r' , $1 \leq r' \neq r \leq \hat{F}$ that satisfies inequality (14),

$$N \sum_{i=1}^{r'} \binom{Q+P}{i} - N \sum_{j=r'+1}^{\hat{F}} H_j \binom{Q+P}{j} \leq f(W^*) \leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r'+1} \quad (15)$$

then, $1 \leq r' \neq r \leq \hat{F} \Rightarrow \hat{F} \geq r' \geq r + 1$ or $r' + 1 \leq r \leq \hat{F}$.

if $\hat{F} \geq r' \geq r + 1$, from the right part of (14),

$$\begin{aligned}
 f(W^*) &= N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r+1} \leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r'} \\
 &= N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - N \sum_{j=r'+1}^{\hat{F}} (H_j + 1) \binom{Q+P}{j} \\
 &= N \sum_{i=1}^{r'} \binom{Q+P}{i} - 1 - N \sum_{j=r'+1}^{\hat{F}} H_j \binom{Q+P}{j} \\
 &< N \sum_{i=1}^{r'} \binom{Q+P}{i} - N \sum_{j=r'+1}^{\hat{F}} H_j \binom{Q+P}{j} \leq f(W^*)
 \end{aligned}$$

This is inconsistent with the fact that $f(W^*) \equiv f(W^*)$.

if $r' + 1 \leq r \leq \hat{F}$, the right part of (15)

$$\begin{aligned}
 f(W^*) &= N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r'+1} \leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_r \\
 &= N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - N \sum_{j=r+1}^{\hat{F}} (H_j + 1) \binom{Q+P}{j} \\
 &= N \sum_{i=1}^r \binom{Q+P}{i} - 1 - N \sum_{j=r+1}^{\hat{F}} H_j \binom{Q+P}{j} \\
 &< N \sum_{i=1}^r \binom{Q+P}{i} - N \sum_{j=r+1}^{\hat{F}} H_j \binom{Q+P}{j} \leq f(W^*)
 \end{aligned}$$

This is inconsistent with the fact that $f(W^*) \equiv f(W^*)$.

Hence, inequality (14) is satisfied by a unique r .

Secondly, if $F^* = r = \hat{F}$, then F^* is the maximum energy well hyper-radius. If $F^* = r < \hat{F}$, then the conclusion that $F^* = r$ is the maximum energy well hyper-radius can be proved by contradiction as follows.

If there is a (W^{**}, M^{**}) pair, with the energy well hyper-radius $F^{**} = e$, $1 \leq r < e \leq \hat{F}$, then,

$$\begin{aligned}
 f(W^*) &\leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_{r+1} \\
 &\leq N \sum_{i=1}^{\hat{F}} \binom{Q+P}{i} - 1 - H_e
 \end{aligned}$$

$$= N \sum_{i=1}^e \binom{Q+P}{i} - H_e + N \sum_{j=e+1}^{\hat{F}} \binom{Q+P}{j} - 1$$

while

$$f(W^{**}) \geq N \sum_{i=1}^e \binom{Q+P}{i} - N \sum_{j=e+1}^{\hat{F}} H_j \binom{Q+P}{j}$$

so,

$$\begin{aligned} f(W^*) - f(W^{**}) &\leq N \sum_{i=1}^e \binom{Q+P}{i} - H_e + N \sum_{j=e+1}^{\hat{F}} \binom{Q+P}{j} - 1 \\ &\quad - \left[N \sum_{i=1}^e \binom{Q+P}{i} - N \sum_{j=e+1}^{\hat{F}} H_j \binom{Q+P}{j} \right] \\ &= N \sum_{j=e+1}^{\hat{F}} (H_j + 1) \binom{Q+P}{j} - H_e - 1 \\ &= -1 < 0 \end{aligned}$$

Then we obtain $f(W^{**}) > f(W^*)$ which is inconsistent with equation (13) that defines W^* as the optimal solution. So F^* is the maximum energy well hyper-radius.

Thirdly, since F^* is the maximum energy well hyper-radius, for any $F' > F^*$, there is at least one neighbor pair $(X, Y) \in \bigcup_{i=1}^N \Omega(X_i, Y_i, n)$, $F^* + 1 \leq n \leq F'$ which has energy lower than or equal to that of one pair $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$. Then if this neighbor pair X', Y' is input to BAM, the output pair will not be the correct training pair. Since $\bigcup_{i=1}^N V_i(F' - 1, M) = \bigcup_{i=1}^N \left[\bigcup_{j=0}^{F'-1} \Omega(X_i, Y_i, j) \right]$ and $(X', Y') \in \Omega(X, Y, 1) \cap \left[\bigcup_{i=1}^N \Omega(X_i, Y_i, n-1) \right]$, $F^* + 1 \leq n \leq F'$, we can obtain that $(X', Y') \in \bigcup_{i=1}^N V_i(F' - 1, M)$. So, there is at least one input pair $(X', Y') \in \bigcup_{i=1}^N V_i(F' - 1, M)$, such that if it is input to BAM, the network does not converge to the correct training pair. Hence, the optimum tolerance set is $V^*(M^*) = \bigcup_{i=1}^N V_i(F^* - 1, M^*)$. ■

Remarks: The optimum noise tolerance set $V^*(M^*) = \bigcup_{i=1}^N V_i(F^* - 1, M^*)$ will be called the *maximum noise tolerance set*. It is for a fixed training pair set. It is possible to find some method, such as the dummy augmentation in [2] to change the set of training pairs to one with increased separation between the training pairs but with the same information content. Intuitively, this can

increase the probability of finding a larger maximum noise tolerance set due to an increased energy well hyper-radius upper bound.

There are three types of neighbors for BAM: 1) the ones $\in V^*(M^*)$, whose output pairs converge to the correct training pairs; 2) the ones, whose deviations are beyond the upper bound $\hat{F} = \left\lfloor \frac{1}{2} \min \left(\min_{0 \leq i \neq j \leq N} H_x(X_i, X), \min_{0 \leq i \neq j \leq N} H_y(Y_i, Y) \right) + 1 \right\rfloor$, whose output pairs will not converge to correct training pairs; 3) others that may or may not be recalled correctly.

Since our approach is based on the energy surface, using different energy definitions, it can be applied to obtain max noise tolerance sets for the higher capacity BAM [9]-[12] rather than the basic BAM only.

IV. COMPUTER SIMULATIONS

A numerical example is given in this section to evaluate the performance of the extended BAM with optimized training weights. Suppose one wants to distinguish three pattern pairs shown in Fig. 2. $X_1 = (-1, -1, -1, -1, -1, -1, -1, 1, -1, -1, -1, 1, 1, 1, -1, -1, -1, 1, -1, -1, -1, -1)$

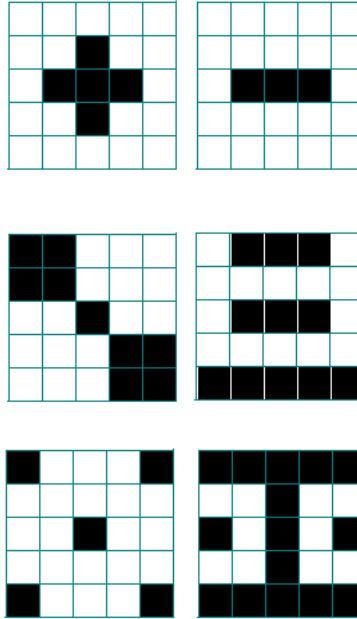


Fig. 2. Three Training Pairs

$$Y_1 = (-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1)$$

$$X_2 = (1,1,-1,-1,-1,1,1,-1,-1,-1,-1,-1,1,-1,-1,-1,-1,1,1,-1,-1,-1,1,1)$$

$$Y_2 = (-1,1,1,1,-1,-1,-1,-1,-1,-1,-1,1,1,1,-1,-1,-1,-1,-1,1,1,1,1,1)$$

$$X_3 = (1,-1,-1,-1,1,-1,-1,-1,-1,-1,-1,-1,1,-1,-1,-1,-1,-1,1,-1,-1,-1,1)$$

$$Y_3 = (1,1,1,1,1,-1,-1,1,-1,-1,1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,1)$$

So,

$$H_x(X_1, X_2) = 12, H_y(Y_1, Y_2) = 8$$

$$H_x(X_1, X_3) = 8, H_y(Y_1, Y_3) = 16$$

$$H_x(X_2, X_3) = 8, H_y(Y_2, Y_3) = 8$$

$$\hat{F} = 8/2 + 1 = 5$$

In this example, to find the optimum training weights, the objective function defined in equation (8) is used as the fitness function of Genetic Algorithm (GA). The results obtained from GA are optimal with high probability. This is acceptable in real applications.

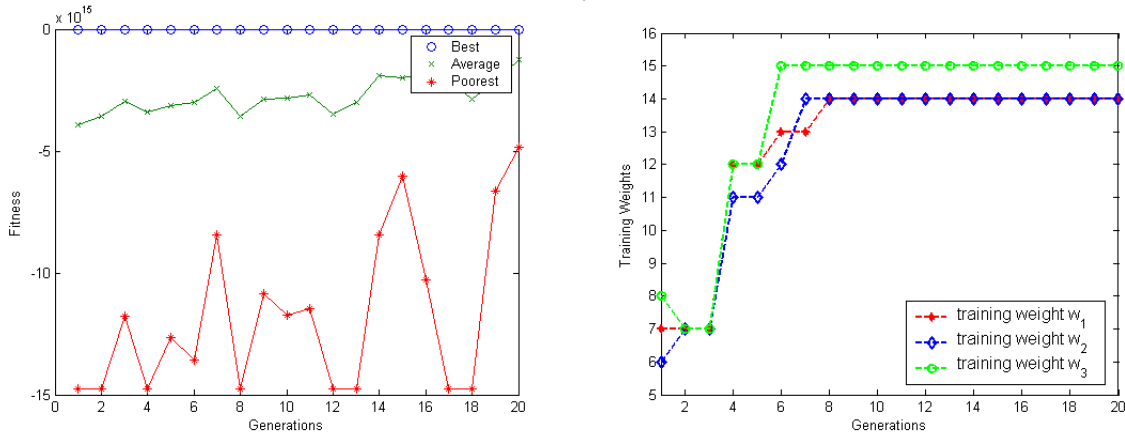


Fig. 3. Fitness Plot and Training Weights

$W^* = (w_1^*, w_2^*, w_3^*) = (14, 14, 15)$, and $F^* = 2$. All training pairs have been recalled correctly and all noisy input pairs with one Hamming distance away from the training pairs have converged to the correct training pair.

V. CONCLUSION

We extended the Basic BAM, using an optimization-based training strategy. For a given set of training pairs, we determined the weights for the training pairs in the BAM correlation matrix

that result in the maximum noise tolerance set. Any noisy input pair within the tolerance set will converge to the correct training pair. We proved that for a given set of training pairs, the maximum noise tolerance set is the largest in the sense that at least one pair, with Hamming distance one larger than the hyper radius associated with the optimum noise tolerance set, will not converge to the correct training pair. A standard Genetic Algorithm (GA) was used to calculate the weights to maximize the object function.

For BAM applications, the speed of encoding is relatively less important than that of the decoding because the encoding can be calculated offline. However, if adaptive encoding is needed to apply to some new desired pairs in real time simulation, the training time should be as short as possible. In the example of this paper, a standard GA algorithm was used. This GA worked well but performed relatively inefficiently, as calculation times were quite long with many generations and fitness values needed to find the optimal solution. Improving the performance of the BAM weight optimization is another future research direction.

ACKNOWLEDGMENT

This research was sponsored by the Defense Advanced Research Project Agency (DARPA) under Contract F33615-01-C3151 issued by the AFRL/VAK. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the DARPA or AFRL.

The authors acknowledge helpful discussions with Dr. Genshe Chen.

REFERENCES

- [1] B. Kosko, "Bidirectional associative memories," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 49–60, Jan. 1988.
- [2] Y.-F. Wang, J. B. Cruz, Jr., and J. H. Mulligan, Jr., "Two coding strategies for bidirectional associative memory," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 81–92, Mar. 1990.
- [3] ———, "Guaranteed recall of all training pairs for bidirectional associative memory," *IEEE Trans. Neural Networks*, vol. 2, no. 6, pp. 559–567, Nov. 1991.
- [4] K. Haines and R. Hecht-Nielsen, "A bam with increased information storage capacity," *Proc. IEEE ICNN 88*, vol. 1, no. 6, pp. 181–190, July 1988.
- [5] R. McEliece, E. Posner, E. Rodemich, and S. Venkatesh, "The capacity of the hopfield associative memory," *IEEE Trans. Inform. Theory*, vol. 33, no. 4, pp. 461–482, July 1987.
- [6] J. Anderson, "A memory storage model utilizing spatial correlation functions," *Kybernetik*, vol. 5, no. 3, pp. 113–119, 1968.

- [7] W. G. Wee, "Generalized inverse approach to adaptive multiclass pattern classification," *IEEE Trans. Comput.*, vol. C-17, no. 2, pp. 1157–1164, Dec. 1969.
- [8] T. Kohonen and M. Ruohonen, "Representation of associative pairs by matrix operations," *IEEE Trans. Comput.*, vol. 22, pp. 701–702, July 1973.
- [9] C. Leung, "Encoding method for bidirectional associative memory using projection on convex sets," *IEEE Trans. Neural Networks*, vol. 4, no. 5, pp. 879–881, Sept. 1993.
- [10] —, "Optimum learning for bidirectional associative memory in the sense of capacity," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 5, pp. 791–796, May 1994.
- [11] B. Lenze, "Improving leung's bidirectional learning rule for associative memories," *IEEE Trans. Neural Networks*, vol. 12, no. 5, pp. 1222–1226, Sept. 2001.
- [12] C.-C. Wang and H.-S. Don, "An analysis of high-capacity discrete exponential bam," *IEEE Trans. Neural Networks*, vol. 6, no. 2, pp. 492–496, Mar. 1995.